DOCUMENT RESUME

ED 441 022 TM 030 820

AUTHOR Johanson, George A.; Osborn, Cynthia J.
TITLE Acquiescence as Differential Person Functioning.
PUB DATE 2000-04-00
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Counselors; Reliability; *Responses; *Surveys; *Test Items; Test Results
IDENTIFIERS *Acquiescence

ABSTRACT

The tendency for some survey respondents to be unreasonably agreeable on attitude measures with an agree-disagree format is typically termed acquiescence, or acquiescent responding (AR). This paper suggests an alternative operational definition of AR plus a statistical test for identifying AR subjects. If positively and negatively phrased items are included in attitude scales to balance the effect of AR, then traditional methods of measuring AR often require computing the difference between the sum of responses to positive items and the sum of (reverse scored) responses to negative items. Large differences are an indication of more AR. Differential person functioning (DPF) can be determined, and AR can be defined as statistically significant DPF between positively and negatively worded item groups. The Theoretical Orientation Scale for Clinicians (TOSC), an inventory of principles of a new therapeutic approach known as solution-focused brief therapy, was completed by 284 counselors (175 usable inventories completed). Thirty-two subjects were identified as "yea sayers," and 8 were identified as "nay sayers." Removing these 40 DPF subjects yielded somewhat improved reliability and factor structure for the scale. The implications of removing DPF subjects from analysis are discussed. (Contains 2 figures and 19 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 022

Acquiescence as Differential Person Functioning

George A. Johanson, Ohio University
Cynthia J. Osborn, Kent State University

A Paper Presented at the 2000 Annual Meeting of the
American Educational Research Association, New Orleans, LA

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

G. Johanson

Please send requests for copies to: George Johanson
201 McCracken Hall
Ohio University
Athens, OH 45701
or
johanson@ohio edu

BEST COPY AVAILABLE

TM030820

The well-known tendency for some survey respondents to be unreasonably agreeable on attitude measures with an agree-disagree response format is typically termed *acquiescence* or *acquiescent responding* (AR). Acquiescence has been of interest for a long time (Lentz,1938), has been occasionally controversial (Rorer, 1965; McGee, 1962), and can threaten the validity of survey measures. Many authors recommend that both favorably and unfavorably phrased items be used in attitude scales with an agree-disagree format to reduce the import of AR (e.g., Best & Kahn, 1998; Crowl, 1996).

The purpose of this paper is to suggest an alternative operational definition of AR plus a statistical test for identifying AR subjects and to illustrate the benefits of detection, particularly in the areas of attitude scale construction and evaluation.

Acquiescence

Those who practice positive AR are frequently referred to as 'yea-sayers' while those who tend to be more or less consistently disagreeable are the 'nay-sayers'. The professional literature is not uniform as to whether AR is a response *style*, *set*, *trait*, or something else entirely. There is considerable literature regarding the consistency and correlates of AR (see, for example, Krosnick, 1999). The notion of *satisficing* (Krosnick, 1991) is appealing and implies that some subjects might tend to agree with any item that seems reasonable in an effort to minimally satisfy the demands of the questioner. Our concern at the moment, however, is less with the various theories or conceptualizations of AR and more with detection and management.

Method

If we include both positively and negatively phrased items in our attitude scales to balance the effect of AR (and also lessen the impression of researcher bias), then traditional approaches to measuring acquiescence often require computing the difference between the sum of responses to positive items and the sum of (reverse-scored) responses to negative items (e.g., Davison & Srichantra, 1988). Larger differences are an indication of more AR.

Differential Functioning

When an achievement test item is easier for one group of examinees than it is for another, it is typically referred to as *item impact* (Dorans, 1989) and may or may not be sensible. However, an item in an achievement test functions differentially for two groups of persons if the item is easier for one group than the other group *after controlling for an overall measure of person skill* (Dorans & Holland, 1993). Such item behavior is referred to as *differential item functioning* (DIF). In an attitude scale, DIF means that an item is easier to agree with for one group of respondents than another after conditioning on a measure of overall person attitude (Johanson, 1997).

If the usual person-item data matrix is transposed to an item-person matrix, then we can determine if a person is functioning differentially between two groups of items after controlling (or accounting) for some overall measure of item agreement. This can be referred to as *differential person functioning* (DPF) and many of the methods of DIF detection can be used with the transposed matrix (Johanson & Alsmadi, 1997).

DPF Detection

There are a variety of empirical methods to detect DIF or DPF (Camilli & Shepard, 1994). Of the classical methods, the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959; Dorans, 1989) is well known and often recommended (Holland & Thayer, 1988; Dorans & Holland, 1993) for DIF detection with binary items. The MH essentially combines 2×2 frequency tables (agree-disagree response x item phrasing) over levels of a third (conditioning) variable into an approximate χ^2 test statistic with one degree of freedom. The null hypothesis

BEST COPY AVAILABLE

tested is whether the ratio of the odds of agreeing with an item from the first group to agreeing with an item from the second group is unity. That is, the null of no differential functioning using the MH method is H_0: odds-ratio=1.

Detecting differential functioning with a transposed data matrix means that the sample size is the number of items in the scale. While the MH procedure does not necessarily require large sample sizes (Allalouf, et al., 1999), it is true that only those individuals with more extreme DPF will consistently be detected as statistically significant when sample sizes are small. Dorans & Holland (1993) state that the Educational Testing Service defines a 'large' effect for DIF be one in which the MH test is statistically significant and where the absolute value of 2.35 times the natural logarithm of the odds ratio be at least 1.5.

Acquiescence as DPF

Person impact might be used to describe a person who simply agrees differently with, say, positively and negatively phrased items since this is precisely what is meant by *item impact* in the original person-item matrix. However, when the same person-difference across item groups is conditioned by some overall measure of item agreement, then this would more accurately be referred to as DPF between positively and negatively phrased item groups and attributable to AR. Our suggestion is that AR, in fact, be operationally defined as statistically significant DPF between these item groups. The question of interest is whether identifying persons showing DPF is actually advantageous in the process of scale construction.

Example: The Theoretical Orientation Scale for Clinicians

The Theoretical Orientation Scale for Clinicians (TOSC) is a 40-item pencil-and-paper, self-report inventory specifically designed to identify the principles of a fairly new therapeutic approach known as solution-focused brief therapy (SFBT). The TOSC is also intended to simultaneously assess one's level of endorsement of such principles. Eleven assumptions of SFBT culled from the literature comprise the theoretical underpinnings of the TOSC and six mental health professionals with expertise in SFBT critiqued and contributed to an earlier version of the instrument. TOSC items are in the form of statements, 15 items are negatively phrased, and response selection is based on a forced-choice 4-point Likert scale (4 = "Strongly Agree," 3 = "Agree," 2 = "Disagree," and 1 = "Strongly Disagree").

The TOSC was initially completed by a random sample of 284 members of the National Association of Alcoholism and Drug Abuse Counselors who responded (a 63% response rate) to a mailed questionnaire. Only those returned questionnaires with 10 or fewer missing observations on the TOSC were considered. Nine cases contained 11 or more missing observations and these were discarded. Of the remaining 275 cases, 63 contained 10 or fewer missing observations. Item-level mean substitution was used for these cases.

The mean age of respondents was 48 years (age range = 23 to 79 years), and 60% were female. Ethnic identity was primarily Caucasian (88%), with African American ranking second (6%), followed by Native American (3%) and Hispanic/Latino (2%). The majority (72%) of respondents reported having earned at least a Bachelor's degree. Of these, 44% stated they held a Master's degree and 6% had earned a Doctoral degree. The majority (83%) of respondents indicated they were certified as alcoholism/drug counselors, and 59% reported working with out-patients, while 41% of respondents stated they worked in private facilities. Licensed social workers comprised the largest professional group (15%), followed by licensed professional counselors (9%), and registered/licensed professional nurses (4%).

Results

The 40 items on the TOSC were put into five groups of relatively homogeneous levels of agreement using quintiles of a binary recoding (1=agreement, 0=disagreement) of responses. The variable identifying these five groups was used for conditioning where there was, over all persons, the least agreement with items in group one and the most agreement with items in group five. Thirty-two (12% of N=275) subjects were identified as 'yea-sayers' or having statistically significant (α=.05) differential functioning using the MH procedure. The plot of the responses to an illustrative 'yea-sayer', person A, is shown in Figure 1. Person A agreed with 80% of the

<insert Figure 1 about here>

positively phrased items in the second group of items, but *agreed with* (the italics remind us that the negatively phrased items have been reverse-scored and that the actual or original response was 'disagreement') only 33% of the negatively phrased items. A person not responding differentially would be expected to have similar levels of agreement (and, thus, coincidental lines for this type of plot). This pattern of more agreement with both positively and negatively phrased items is consistent across the first four item-agreement groups. Person A agreed with all items in the fifth group of most agreeable items. This is a statistically significant pattern of responding (χ^2=4.058, df=1, N=40 [items], p=.044) with an associated odds-ratio estimate of 9.

Eight subjects (3%) were found to be statistically significant 'nay-sayers'. An example of one such respondent is person B (χ^2=4.242, df=1, N=40, p=.039) in Figure 2. The total number

<insert Figure 2 about here>

of respondents showing some form of AR or DPF was 40 (15%) and all effects were 'large'.

Data for the TOSC were reanalyzed with these 40 subjects removed (N=235). The item analysis was similar to that with N=275, but the reliability (Cronbach's alpha with 40 items) increased slightly from .79 to .82 with the removal of the 40 DPF subjects. The original principal components analysis (N=275) showed a somewhat suspect factor structure with the second of four components reflecting mainly negative phrasing (11 of the 12 items loading at .3 or greater on this factor were negatively phrased). The factor structure with N=235 was more appropriate with three factors retained and with the absence of a troublesome 'negative phrasing' factor. With N=275, the correlation between the sum of responses to the (15) negative items and the sum of the (25) positive items was essentially zero (r=.050, p>.05) while with N=235, the correlation was a more reasonable r=.323 (p<.01). In short, the scale was found to have somewhat improved reliability and factor structure when the DPF respondents were removed.

Discussion

Should subjects be removed from analyses simply because they have responded to survey items in a manner the researcher finds unreasonable? One position is that acquiescent respondents will have little or no effect on estimates of key parameters (means) if the scale has approximately equal numbers of positively and negatively phrased items (e.g., Mueller, 1986; Spector, 1992). Krathwohl suggests that both positive and negative phrasing can be used for each item and then the researcher can "...eliminate the responses of people who contradict themselves." (p. 392). He goes on to caution, however, that removing respondents may adversely effect the generality of the study.

BEST COPY AVAILABLE

While these assertions are certainly sensible, another position would be to identify respondents with a significant amount of DPF and remove them from scale development analyses by appealing to the same rationale that is used when removing items with DIF from an achievement test. That is to say, when an item on an achievement test is found to be functioning differentially and inappropriately favors one group over another (i.e., biased), the item is almost always removed for subsequent (person) analyses. The same logic, of course, would imply that persons found to be functioning differentially and substantially favoring one item type over another (acquiescing) should be removed for subsequent item analyses. It is comforting to note that the prevalence of AR we noted in our example (12%) is not far from the estimate of 10% recently reported in the literature by Krosnick (1999) across a variety of studies and measures.

Current methods of identifying AR are limited in that they correspond to the notion of 'impact' and do not come with a corresponding statistical test. Item impact and DIF can be quite different in achievement testing. If one group of students has been instructed and another not, then relevant achievement items will likely show evidence of impact. Simple (or unconditioned) group differences in performance can be desirable. DIF, on the other hand, is a different and more serious problem because DIF implies that persons who are similar in overall achievement still differ on an item and, thus, the difference must be due to something other than achievement. We contend that the same is true of DPF with positively and negatively phrased item formats and would label this 'something' *acquiescence*.

References

Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. Journal of Educational Measurement, 36(3), 185-198.

Best, J. W., & Kahn, J. V. (1998). Research in education (8th ed.). Needham Heights, MA: Allyn & Bacon.

Camilli, G., & Shepard, L. A. (1994). Methods for identifying biased test items. Thousand Oaks, CA: Sage Publications.

Crowl, T. K. (1996). Fundamentals of educational research (2nd ed.). Madison, WI: Brown & Benchmark.

Davison, M. L., & Srichantra, N. (1988). Acquiescence in components analysis and multidimensional scaling of self-rating items. Applied Psychological Measurement, 12(4), 339-351.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. Applied Measurement in Education, 2(3), 217-233.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland, & H. Wainer (Eds.), Differential item functioning (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer, & H. I. Braun (Eds.), Test validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Johanson, G. A. (1997). Differential item functioning in attitude assessment. Evaluation Practice, 18(2), 127-135.

Johanson, G. A. & Alsmadi, A. (1998, April). Differential person functioning. A paper presented at the annual meeting of the American Educational Research Association in San Diego, CA.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measurement in surveys. Applied Cognitive Psychology, 5, 213-236.

Krosnick, J. A. (1999). Survey research. Annual review of psychology, 50, 537-567.

Krathwohl, D. R. (1993). Methods of educational and social science research: An integrated approach. New York: Longman.

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality [Abstract]. Psychological Bulletin, 35, 659.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

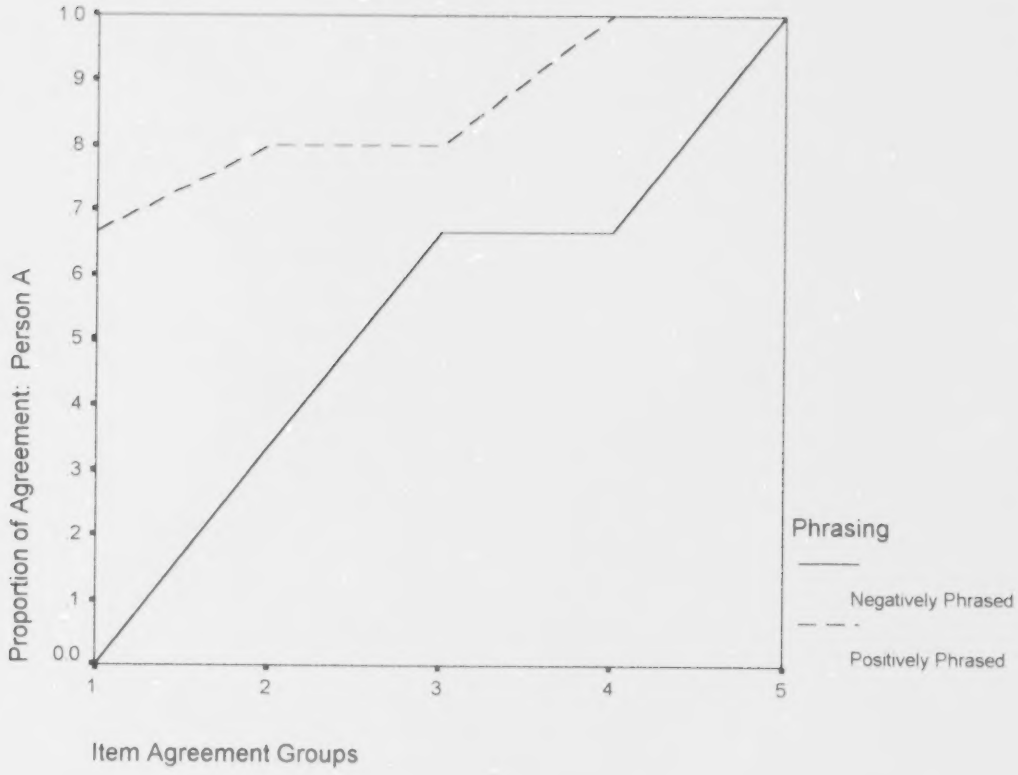
McGee, R. K. (1962). Response style as a personality variable: By what criterion? Psychological Bulletin, 59(4), 284-295.

Mueller, D. J. (1986). Measuring social attitudes: A handbook for researchers and practitioners. New York: Teachers College Press.

Rorer, L. G. (1965). The great response-style myth. Psychological Bulletin, 63(3), 129-156.

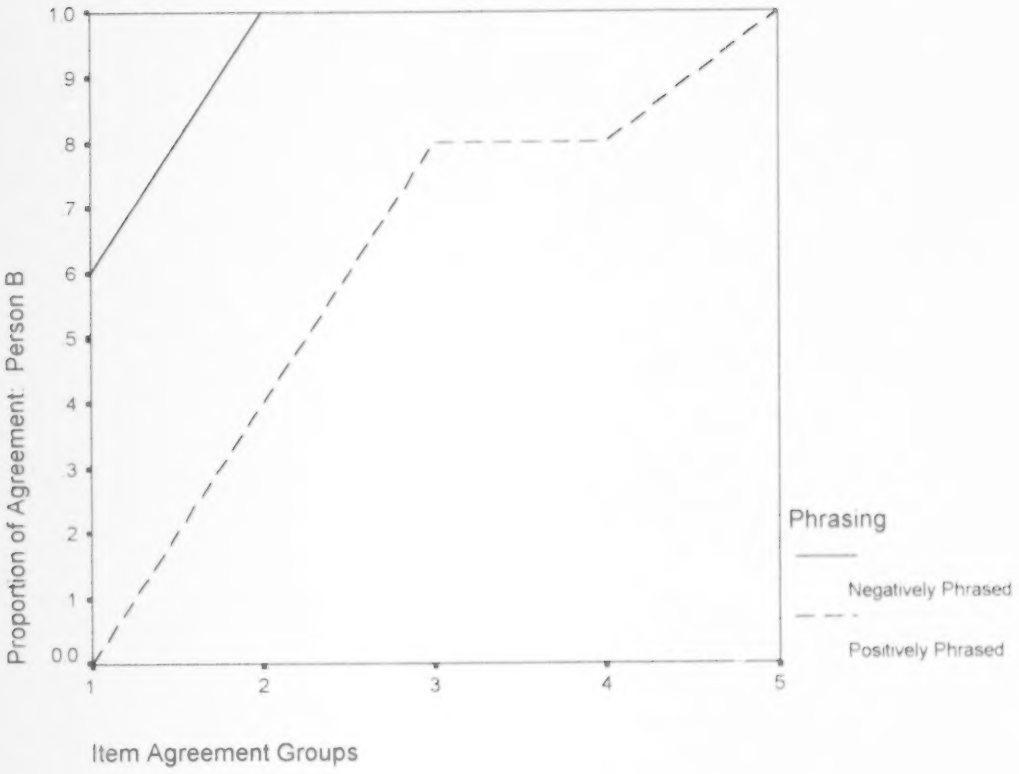
Spector, P. E. (1992). Summated rating scale construction: An introduction. Newbury Park, CA: Sage Publications.

Figure 1. An Example of a 'Yea-Sayer'.



BEST COPY AVAILABLE

Figure 2. An Example of a 'Nay-sayer'.



BEST COPY AVAILABLE